bio-protocol Preprint

Not Peer Reviewed
Ask a question online

SUrface-protein Glycan And RNA-seq

Conor J. Kearney Jane Oliaro Ricky W. Johnstone

Updated date: Jul 5, 2021

An abbreviated version of this protocol was published in Science Advances in Feb 2021

SUGAR-seq enables simultaneous detection of glycans epitopes and the transcriptome in single cells

DOI: 10.1126/sciadv.abe3610

Detailed protocol

SUGAR-seq protocol for tumor infiltrating lymphocytes (TILs)

FACS-sorted TILs were incubated with a mixture of biotinylated and non-biotinylated L-Pha (1:5 ratio), to prevent signal saturation, at a concentration of 1 μg/ml for 20 minutes on ice. After three washes with cold phosphate-buffered saline (PBS), cells were stained with cell hashing antibodies and CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) antibodies as previously described (*1*). An anti-biotin oligo-conjugated CITE-seq antibody was used to convert the L-Pha signal into a sequenceable readout. “Stained” and washed cells were counted, brought to ~1000 cells/μl, and loaded onto the 10× chromium instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell gel beads in emulsion and capture/barcode cells. For 5′ multiomics including gene expression, TCR-sequencing and hashing/CITE-seq, we used the 10× Single Cell V(D)J Kit with Feature Barcoding (enabled for 5′ gene expression, TCR, and feature barcoding for cell surface protein) protocol following the manufacturer’s instructions. TCR libraries were prepared using the 10x Chromium Single Cell V(D)J Enrichment Kit, mouse T cell. 5′ hashtag oligos (HTO)/ADT (one library) was prepared using the Chromium Single Cell 5′ Feature Barcode Library Kit and indexed using the Chromium i7 Multiplex Kit N, Set A. Cell Ranger reads were aligned to the mm10 reference genome; cellular barcodes were demultiplexed, and unique molecular identifiers and antibody capture (ADT and HTO) were quantified using 10x Genomics’ Cell Ranger software (version 3.1.0). Cell barcodes containing RNA or antibody counts from cells from more than one sample (intersample doublets) were identified using Seurat’s HTODemux function. Barcodes containing counts from more than one cell within the same sample (intrasample doublets) were identified using the Scrublet Python package (version 0.2.1) (*2*). A cutoff of more than one median absolute deviation value above the median Scrublet score was chosen. Cells identified as either type of doublet were removed from the analysis. Relative TF activity in each cell was estimated using the single-cell regulatory network inference and clustering method (*3*). Area-under-curve (AUC) scores for TFs were calculated on the basis of inferred gene regulatory networks using the pySCENIC Python package (version 0.9.19). Gene expression and antibody count matrices were processed in R (version 3.6.1) using the Seurat R package (version 3.1.0) (*4*).RNA transcript counts for barcodes identified as cells by Cell Ranger were normalized using sctransform via Seurat’s SCTransform function. The counts were first transformed with no covariates in the sctransform model; then, cell cycle phase scores were estimated using Seurat’s CellCycleScoring function with mouse homologs of the cell cycle gene sets provided by Seurat. The sctransform normalization was then rerun with the cell cycle phase scores and the percentage of raw RNA counts belonging to mitochondrial genes for each cell as variables to be regressed out in the model. ADT counts were log1p transformed and scaled (zscore) or normalized using centered log ratio transformation with either method, producing a similar result. Principal components analysis was then performed on the sctransform-scaled RNA expression values for genes with residual variance in the sctransform model greater than 1.3. A shared nearest neighbor (SNN) network was calculated using the top 10 principal components using the FindNeighbors function with *k*-nearest neighbors set to 50 and cosine distance metric. The SNN network was then used to identify cell populations using the FindClusters function using the Louvain algorithm with resolution parameter 0.6. UMAP values were also calculated using the RunUMAP function with the top 10 principal components as input and parameters n.neighbors = 50 and metric = “cosine.” Monocle3 was used for pseudotime analysis using default setting.

References

1. M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868 (2017).
2. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289 (2019).
3. S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van der Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
4. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Kearney, C. , Oliaro, J. and Johnstone, R. (2021). SUrface-protein Glycan And RNA-seq. Bio-protocol Preprint. bio-protocol.org/prep1257.
2. Kearney, C. J., Vervoort, S. J., Ramsbottom, K. M., Todorovski, I., Lelliott, E. J., Zethoven, M., Pijpers, L., Martin, B. P., Semple, T., Martelotto, L., Trapani, J. A., Parish, I. A., Scott, N. E., Oliaro, J. and Johnstone, R. W.(2021). SUGAR-seq enables simultaneous detection of glycans epitopes and the transcriptome in single cells . Science Advances 7(8). DOI: 10.1126/sciadv.abe3610